DOCUMENT RESUME

ED 235 194 TM 830 586

AUTHOR Furst, Edward J.
TITLE Communicability of the Taxonomy of Educational Objectives for the Cognitive Domain.
PUB DATE Jun 83
NOTE 26p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; *Classification; Cognitive Processes; *Diffusion (Communication); *Educational Objectives; Evaluation Methods; Instructional Materials; *Test Items; Test Validity
IDENTIFIERS *Blooms Taxonomy; *Communicability

ABSTRACT

Enough evidence has accumulated on Bloom's "Taxonomy of Educational Objectives" for the cognitive domain to justify a review of its communicability. This article covers both published and unpublished studies as well as certain informal reports that bear on this property. It also examines possibilities for improving agreement among educators using this scheme or minor revisions. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235194

Communicability of the Taxonomy of Educational Objectives for the Cognitive Domain

Edward J. Furst

University of Arkansas

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. J Furst

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 586

Running Head: Communicability of Taxonomy

Abstract

Enough evidence has accumulated on the Taxonomy of Educational Objectives for the cognitive domain to justify a review of its communicability. This article covers both published and unpublished studies as well as certain informal reports that bear on this property. It also examines possibilities for improving agreement among educators using this scheme or minor revisions.

Communicability of the Taxonomy of Educational Objectives for the Cognitive Domain[1]

Extensive use has been made of the Taxonomy of Educational Objectives for the cognitive domain (Bloom, et al., 1956), here called simply the taxonomy. Despite this wide use such issues as communicability and the assumption of a cumulative hierarchy still remain. As the literature on either of these properties is now large, this article will address only the first. The author will attempt a review and critique of the evidence on communicability, a property important in its own right because prerequisite to the usefulness of the scheme.

Bloom et al. (1956) refer to communicability essentially as the ease with which educators agree in classifying objectives and test materials; so it is not surprising that most studies of the problem have focused on those matters. An occasional few have dealt with the classification of oral questions during teaching or of objectives and mental operations inferred from instructional plans. These extensions appear to be in keeping with the suggested uses of the taxonomy.

It may seem reasonable to extend the notion of communicability to include agreement among specialists who use statements of objectives as specifications for developing test materials, instructional materials, and instructional activities. At issue is the question whether different specialists, working from the same specifications, would produce very similar products. For this to happen would require prior agreement on theory relevant to the particular kind of activity. That is to say, all three of these developmental activities go beyond classification per se and invoke or require theory. The mere statement of educational objectives is not sufficient to generate rules

for guiding such activities. Moreover, the required theory may not yet have been worked out, let alone validated and accepted. A case in point is item-writing theory and item-response theory, which relate to the development of test items. Bormuth's contention (1970, pp. 66-83) that we have no adequate theories for these two domains is probably still true today. Thus to extend study of the communicability of the taxonomy to such developmental activities would seem somewhat premature.

The Question of What is to be Classified

Having delimited communicability as above, and considering also the title of the handbook, it may appear that the question of what is to be classified has been well answered. The focus is to be educational objectives. It is a fact, however, that as much as half the handbook is given over to illustrative test items (using this term broadly to include exercises, problems, questions, and choice-response items); also, that most studies of communicability have concentrated on having judges classify test items rather than statements of objectives. In view of the prominence given to test items and the implied question as to how test items stand in relation to objectives, we must take a close look at how definition was carried out in the taxonomy.

The plan was to carry out definition on three levels (Bloom, et al., 1956, pp. 44-45):

I. Verbal description of each class and subclass

II. Listing of statements of particular objectives under each subclass

III. Illustrations of test exercises, problems, questions, and items

At the first level, verbal definition can go only so far in conveying meaning about the abstract constructs named, for the terms carry different meanings in different contexts and for different persons. Hence the authors had to resort to ostensive and operational definitions at both the second and third levels to avoid this impasse (Wilhoyte, 1965, pp. 37-47). Even so, definition was at best approximate because educational objectives in the cognitive domain refer ultimately to intended changes in students' thinking--to mental acts, which are not directly observable.

> Type III definitions do not define educational objectives directly. Each test item functions as an ostensive definition of the class characteristic of which it is an instance. Test items also function as operational definitions insofar as it is understood that what is defined (student behavior) is not classes of test items, but classes or types of responses to test items...(Wilhoyte, 1965, p. 46).

The foregoing remarks suggest why test items have played so large a role in explicating the meaning of the taxonomy and why researchers have concentrated on them when assessing its communicability with educators. This approach to operational definition has limitations, some of which will be brought out later. A recent review of philosophical criticism on this and related problems with the taxonomy may be found in Furst (1981). With this background let us turn now to a review of available evidence.

Studies of Communicability

As noted, most studies of communicability of the taxonomy have assessed the agreement among educators in classifying objectives or

test items according to the different categories, as follows:

1.00 Knowledge	4.00 Analysis
2.00 Comprehension	5.00 Synthesis
3.00 Application	6.00 Evaluation

Except for a very few studies, judges have been asked to classify only by these major categories and not to the level of the subcategories.

In a comprehensive review of research on educational and psychological issues concerning the taxonomy, Seddon (1978) touched only briefly on communicability. He located eight studies on classification of test items by major taxonomic levels and noted a trend for percent agreement to decrease as the number of judges increases. As is also true of later studies to be reviewed here, it was not possible to determine what other factors might have accounted for the differences among the several investigations. Seddon did generalize, however, that a majority of those researchers were too sanguine in their conclusions. He regarded their conclusions as misleading in light of the findings of Fairbrother (1975) whose judges (teachers of physics) supposedly were more representative of a meaningful population of educators. Yet the report of Fairbrother's study reveals that he worked with judges untrained in the use of the taxonomy. Each volunteer received about a two-to-three-page leaflet giving a sketch of four levels of the taxonomy (Analysis and Evaluation combined) and including only four illustrative test items--one for each level covered. A covering letter gave brief instructions for doing the classifying. In fairness it must be said that some other studies of categorizing items have not

required much training either. The necessity of having judges well-trained in the task is so basic that it strikes one that not much significance should be given to studies of classifying where so little training has been given.

Of the hundreds of studies on the cognitive domain, we were able to locate several in addition to Seddon's list that offered some firm data on agreement and that also had required judges to undergo at least minimal training. These several studies appear in Table 1. An inspection will show that it is very difficult to compare any two

Insert Table 1 About Here

of these on relative agreement because of differences in what is classified, the subject matter, level of instruction, and other conditions. In one sense, the results seem quite favorable since (except for the use of the coefficient of correlation) they indicate extent of agreement under the most severe standard--i.e., placing items in the same category of the taxonomy (or in the case of Klein, in the same subcategory). In another sense, however, the evidence is misleading because the task had often diminished to a few levels, either by default as when a preponderance of the items were on Knowledge and Comprehension (e.g., Agishi, 1976; Pancella, 1970) or by design as when the investigator concentrated on three or four levels (e.g., Little, 1971, and Unruh, 1974, levels 1.00 to 4.00).

A recent study (Baldwin, 1980) can be included here even though it gave no training to judges. Baldwin required that participants have "previous exposure" to Bloom's taxonomy, that English be their

native language, and that they affirm their ability to classify a set of test items using the taxonomy. Drawing on two populations, graduate students in instructional systems at a midwestern state university and faculty/staff at a southern state university, he assigned four treatments randomly and gave them individually. Treatments varied in information about "previous student learning." Judges classified 50 test items or a reading passage (The Lisbon Earthquake, from Kropp & Stoker, 1966). N's, after some small attrition, were 11, 13, 8, and 12, respectively. Coefficients of concordance (Kendall's W) for the classifications within each treatment group were .59, .57, .58, and .62, after dropping an outlier from the second group. Mean Spearman rank-difference coefficients between the judges and the "instructional developer" (Kropp & Stroker, 1966) were .59, .59, .72, and .71. A cross-tabulation of the judges' classifications against those of the instructional developer revealed percents of "errors" by categories, Knowledge through Evaluation, of 34, 58, 75, 73, 42, and 38. Items keyed as Knowledge, Synthesis and Evaluation were thus less often misclassified than those in Comprehension, Application, and Analysis. The two non-adjacent categories of Comprehension and Analysis showed the greatest confusion, 20 percent of the items in the first classified in Analysis, and 29 percent of the latter classified in Comprehension. Considering all the evidence, including the moderately high and statistically significant indexes of correlation, Baldwin concluded that the lack of practical agreement among judges casts some doubt on the utility of the taxonomy as an aid to communication between educators.

This rather negative conclusion must, however, be tempered by certain limitations of the study. By design the first two sets of

judges worked without knowledge of the instructional materials (the reading passage) which would have helped them judge what information the student had access to when attempting the test items. For the first two groups, then, Baldwin's assumption that judges were doing a task analysis on each item does not seem justified because an operational definition of a task must include the givens, such as stimulus materials and directions. Perhaps more important was the uncertainty about the competence of the judges as there was no check on the extent of training and experience with the taxonomy.

Rather more favorable conclusions on communicability were reached by Steele (1969) in a dissertation although they must be qualified because he used a revised version of the taxonomy. Steele divided the second category, Comprehension, into two major categories, Translation and Interpretation/Extrapolation, redefined part of Analysis as "Formal Analysis" and placed it at the highest level because he considered it to represent an extremely rigorous and complex operation. This yielded seven major categories renamed the Taxonomy of Intellectual Abilities (TIA), which was then field-tested several times, together with a set of sample objectives and another of test questions, to improve the descriptions and illustrations. Forty-four experienced public school administrators and teachers taking a graduate course on instructional supervision then provided a test of classification. After getting a brief explanation of the TIA and the larger study, they used the rest of a class period to classify a set of 20 practice objectives and a set of 20 test questions. Each participant devoted from one to two hours out of class to classify about 60 additional objectives and about 90 test questions, each set exclusively

pertaining to an introductory undergraduate course in special education. A majority of these essentially untrained judges agreed on the classification of about two-thirds of these objectives and similarly of the test questions. More significant was the fact that a much higher degree of agreement was reached by the six instructors of special education taking part in the study (and also by six professors who had assisted in field-testing). Two-thirds or more of the instructors agreed on 56 percent of the objectives, and half agreed upon an additional 21 percent, yielding a total of 77 percent thus readily identified with a category. Two-thirds or more agreed on 88 percent of the test questions. Still higher agreement among all groups of judges was achieved when classifications were dichotomized according to lower (I-III) or higher (IV-VII) levels of the TIA. From these and related findings, Steele concluded that the TIA could be used reliably and with precision by teachers and administrators to classify objectives and questions.

Informal Evidence on Communicability

Potentially there is a large body of experience bearing on the issue of communicability, much of it rather subjective and practically inaccessible. Some of this would be favorable to the use of the taxonomy and some unfavorable. An instance of where we do at least have a brief public report of a large-scale trial is the effort in a three-county area of metropolitan Portland, Oregon (Doherty & Hathaway, 1972). The evaluators in that consortium reported that teachers had found it hard to deal with cognitive goals within the context of the taxonomy because it was not clear how the broad descriptors of psychological processes could be treated instructionally. The Portland group

therefore worked out their own classification to ensure communicability with teachers; however, they did acknowledge that the taxonomy was helpful as a starting point.

Mathematics may be of special significance in connection with the question of communicability, for it is a formal subject in which objectives and tasks should be amenable to precise classification. Some indirect and informal evidence has appeared in a few scattered accounts. One of these (Romberg, 1968, pp. 5-10) grew out of the National Longitudinal Study of Mathematical Abilities (NLSMA). At a special conference a group of 39 mathematicians and math educators attempted to classify various achievement test items using the taxonomy as the point of departure. No information was given on the extent of their training, or on their agreement other than that they were unable to categorize items into the taxonomy. Romberg reported that these professionals liked the idea of levels of complexity and that they did outline a similar scheme of seven levels--Knowing, Translating, Manipulating, Choosing, Analyzing, Synthesizing, and Evaluating--a scheme later revised into a model of five general mental processes in mathematics.

In a critical review of the taxonomy, Ormell (1974) used mathematics as a test case because of its precision. He argued from an analysis of mathematical problem solving required in the schools and from experience with the taxonomy that its implementation in that subject is extremely difficult-- an especial source of difficulty in mathematics is that the boundaries between Knowledge, Comprehension, and Application are exceedingly blurred. Few problems in the upper school (secondary) can be uniquely placed in these categories. Moreover, there are a great many "adjustment" steps in math that need to

be taken with great skill, and they can be regarded either as examples (i.e., Application) of general rules or as examples of Comprehension. Because Ormell found the categories too amorphous in relation to the precision of math, he concluded that to try to implement them--at least at the level of the upper school--is "a continuous exercise in arbitrary choice" (p. 7).

In contrast to Ormell's negative view, other authorities on mathematical education have found the taxonomy a useful scheme (Avital, 1968; Wilson, 1971). Each of these, however, has made one or more major changes for secondary school mathematics. Wilson reduced the major categories to four--Computation, Comprehension, Application, and Analysis--the last level incorporating some operations from Synthesis and Evaluation. This model reflects Wilson's work in test construction for the NLSMA. Avital, with the upper elementary grades and the secondary school in mind, followed the taxonomy closely, dropping only Evaluation because he did not consider it a psychologically distinct process in mathematical performance. He concentrated only on the major headings, as shown below, and clearly distinguished them through definition and example (p. 7).

Thinking Process	Taxonomic Level
recognition, recall	1. Knowledge
algorithmic thinking, generalization	2. Comprehension 3. Application
open search	4. Analysis 5. Synthesis

So the evidence from these few citations is decidely mixed and shades into the larger question of usefulness. Since it is not the purpose o. this article to consider this larger question, the reader may turn to other papers (e.g., Furst, 1981) for further information.

Improving Reliability and Objectivity of Classification

Granted that there is room for improvement in communicability, what steps can be taken in using the taxonomy? Improvement in the context of this review refers primarily to the reliability and objectivity of classification. The psychometric literature refers to these matters as the "reliability" of judgments with respect to such diverse phenomena as personality traits, human performance, verbal statements, and physical stimuli. This literature prescribes several important avenues of improvement.

Two prescriptions pertain to the simplicity and clarity of the scheme judges are to use. The categories --in this instance, the major levels of the taxonomy--should be relatively few. Higher reliability and agreement may be expected when statements and items are to be sorted into broad rather than narrow categories. But as some of the cited studies have shown, a small number of broad classes will reduce reliability and agreement if judges find the boundaries indistinct and the categories ambiguous. In the case of educational objectives and test items the further requirement of mutually exclusive and unambiguous categories is unrealistic though it may be approximated. Part of the problem of unreliability or disagreement must arise from the fact that the major classes of objectives have little operational meaning. In any event, many revisions of the taxonomy have already been carried

out for the purpose of improving communicability as well as general usefulness. Thus, the very scheme to be examined has itself been transformed.

Looking at the task from the other side--from what is to be classified--the psychometric movement would suggest two other ways of increasing reliability. These two avenues are through the provision of more information and they amount to operationalization and further specification. It has already been noted that, with respect to test items, judges need to know the givens and other circumstances under which the student must work. Judges also need to have common information on the typical student and his instruction, so that they work from the same premises in classifying. Doubtless much disagreement has crept in because judges lacked this information and made their own assumptions.

A few promising steps have already been taken to reduce unreliability owing to uncertainty about variables such as student experience. One of these involves the use of the taxonomy in developing a "cognitive profile" of instructional materials (Scott, 1972). To do this, one would decide the highest or most complex cognitive level at which a student must work to complete the activity satisfactorily. The problem for the analyst is to take account of both the student's past experience and the teacher's technique of instruction. Scott's tentative solution was to devise a coding system to take account of these two kinds of variables. The basic rule was to code using the lowest cognitive behavior clearly required but to add a higher "possibility" coding. For instance, if an activity

might at the least require interpretation (coded 2.20 in the taxonomy) but might be taught in such a way as to require application (coded 3.00), then the final coding would be "2.20, p 3.00." The higher possibility would thus be contingent on the student's having or not having certain necessary knowledge, the kinds of information or directions the teacher gives, etc. In a similar vein Friedman (1974) formulated a few basic rules in coding teachers' oral questions in geometry, taking account again of whether the student had already learned relevant information, had been exposed to it in immediately preceding questions, etc. Although he recognized that classification on the basis of inferred mental processes still has pitfalls, he agreed with Sanders (1966, pp. 8-9) that, for the most part, teachers can classify questions accurately because they know how they have been teaching the subject. This is an important point, for it implies that ultimately the best judge of objectives of test items is the teacher, the one person in the best position to understand the student and the conditions of instruction.

A further possibility to increase reliability or agreement is the algorithm. Horn (cited in DeLandsheere, 1972, p. 109) suggested the following sequence for the taxonomy:

1. Is it possible to distinguish the content from the model of the problem?
   - No → Knowledge
   - Yes ↓
2. Is all the information needed to solve the problem provided?
   - Yes → Comprehension
   - No ↓

3. Must the student bring supplementary information to solve the problem? — Yes → Application
   ↓ No
4. Must the student examine the situation according to definite criteria? — Yes → Analysis
   ↓ No
5. Is there more than one solution to the problem? — Yes → Synthesis

Horn did not include Evaluation, possibly subsuming it under Analysis. In any event, he concluded that well-trained judges can reach a high and satisfactory degree of agreement.

A final important way of improving reliability and agreement hardly needs elaboration but it does need mention again. That is the careful training of judges in the use of the classificatory scheme.

Over-all Critique on Communicability

It is evident from the studies reviewed here and earlier that, with the possible exception of well-specified statements of objectives, near-perfect agreement in classifying objectives and test items into the major categories is not within easy reach. Communicability has practical limits. Statements of objectives, as well as their counterparts in test items, are always to some degree ambiguous, and ever-finer specification cannot eliminate misunderstanding of intentions (MacDonald-Ross, 1973). In this regard, Bloom et al. (1956) noted the indeterminancy that a test item or exercise cannot be classified in any absolute sense. These authors pointed out the necessity of considering the student's experience (pp. 15-16), the context in which

the test problems were used (p. 21), and the learning situations preceding the test (p. 51). Even so, uncertainty still remains because it is possible for students of like experience to use different operations in answering questions or solving problems. It is this possibility that would help account for the lesser agreement in classifying items at the higher levels--Pancella (1970), for instance, found 90% agreement on level 1.00 items but 65% on levels 3.00 to 6.00 combined.

High agreement of itself, of course, would not indicate validity of the taxonomy. Judges may reach relatively high agreement in classification through impressionistic use of the scheme with shared meanings of definitions and examples. It is necessary to show that the mental operations evoked do correspond to those intended by statements of objectives and that those operations collectively form a hierarchy as postulated by the authors.

Considering the practical limits noted above, the evidence on extent of agreement appears more favorable than the earlier review by Seddon (1978). The evidence looks even more favorable when one considers that the various formal studies over the years have, for the most part, suffered from much the same limitations--the use of judges not really expert in the taxonomy and the failure to provide sufficient information on instructional materials, activities, previous student learning, and conditions of testing. Still further agreement may be expected from use of an apparently improved version of the original taxonomy, containing some simplification as well as refinement, and replete with more illustrative objectives and test items (Bloom, Hastings, and Madaus, 1971).

This more favorable view must be tempered by the fact that some of the studies of communicability have been carried out with revisions of the original taxonomy, ordinarily minor but occasionally major. Such revisions tend to indicate some difficulty with one or more of the major categories and their subdivisions. What seems to be needed now is a series of studies examining relative communicability by systematically comparing the taxonomy, preferably the 1971 revision mentioned above, with the use of other schemes of classification. In this way, meeting the competition would be the decisive test.

References

Agishi, A. A taxonomical analysis of the cognitive abilities measured by the English entrance examinations for admission to colleges and universities in Japan (doctoral dissertation, University of California at Los Angeles, 1976). Dissertation Abstracts International, 1976, 37, 3440A. (University Microfilms No. 76-26,876)

Avital, S. M. and Shettleworth, S. J. Objectives for mathematics learning: some ideas for the teacher. Bulletin No. 3, The Ontario Institute for Studies in Education, Toronto, Canada, 1968, 57 pp.

Baldwin, W. L. Agreement among educators classifying test items according to Bloom's taxonomy (doctoral dissertation, Indiana University, 1980). Dissertation Abstracts, 1980, 41, 906A (University Microfilms No. 8020016)

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. Taxonomy of educational objectives, handbook I: cognitive domain. New York: McKay, 1956.

Bloom, B. S., Hastings, J. T., & Madaus, G. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

De Landsheere, V. On defining educational objectives, Evaluation in Education, 1977, 1, 73-190.

Doherty, V. W., & Hathaway, W. E. Goals and objectives in planning and evaluation: a second generation. NCME Measurement in Education (A series of special reports of NCME), 1972, 4 (1).

Ernst, R. E. A taxonomical analysis of selected units of the Hawaii comprehensive musicianship program. (doctoral dissertation, University of Michigan, 1974). Dissertation Abstracts International, 1975, 35, 4584A. (University Microfilms No. 75-682)

Fairbrother, R. W. The reliability of teachers' judgements of the abilities being tested by multiple choice items. Educational Research, 1975, 17, 202-210.

Friedman, M. A system to analyze geometry teachers' questions. Mathematics Teacher, 1974, 75, 709-713.

Furst, E. J. Bloom's Taxonomy of educational objectives for the cognitive domain: philosophical and educational issues. Review of Educational Research. 1981, 51 (4), 441-453.

Klein, M. F. Use of "Taxonomy of educational objectives (cognitive domain)" in constructing tests for primary school pupils. Journal of Experimental Education, 1972, 40, 38-50.

Kropp, R. P., & Stoker, H. W. The construction and validation of tests of the cognitive processes described in the "Taxonomy of educational objectives" (Cooperative Research Project No. 2117). U.S. Office of Education, 1966.

Little, R. A. A taxonomic approach to measuring achievement in mathematics 223--geometry for elementary teachers (doctoral dissertation, Kent State University, 1971). Dissertation Abstracts International, 1972, 32, 6105-61061A. (University Microfilms No. 72-15,945)

Macdonald-Ross, M. Behavioral objectives: A critical review. Instructional Science, 1973, 2, 1-52.

Mills, S. R., Rice, C. T., Berliner, D. C., & Rosseau, E. W. The correspondence between teachers' questions and student answers in classroom discourse. Journal of Experimental Education, 1980, 48, 194-209.

Ormell, C. P. Bloom's taxonomy and the objectives of education. Educational Research, 1974, 17, 3-18.

Pancella, J. R. Use of the "Taxonomy of educational objectives handbook I: cognitive domain" (Bloom) to analyze and describe standardized and commercial tenth grade biology examinations (doctoral dissertation, University of Maryland, 1970). Dissertation Abstracts International, 1971, 31, 3893A. (University Microfilms No. 71-4522)

Romberg, T. A. Derivation of subtests measuring distinct mental processes within the NLSMA algebra achievement test (doctoral dissertation, Stanford University, 1968). Dissertation Abstracts, 1968, 29, 419A. (University Microfilms No. 68-11,341)

Sanders, N. M. Classroom questions: what kinds? New York: Harper and Row, 1966.

Scott, H. V. The taxonomy of educational objectives as a curriculum analysis tool: a solution to some problems encountered while coding activities. Science Education, 1972, 56, 411-415.

Seddon, G. M. The properties of Bloom's taxonomy of educational objectives for the cognitive domain. Review of Educational Research, 1978, 48, 303-323.

Steele, J. M. Things as they are: an evaluation procedure to assess intent and practice in instruction (doctoral dissertation, University of Illinois, 1969). Dissertation Abstracts International 1970, 30, 3371-A. (University Microfilms No. 70-00992)

Unruh, R. P. An investigation of the construct validity of a taxonomy of reading comprehension (doctoral dissertation, University of Houston, 1971). Dissertation Abstracts International, 1975, 35, 3567-3568B. (University Microfilms No. 75-1037)

Wilhoyte, R. L. Problems of meaning and reference in Bloom's taxonomy: cognitive domain (doctoral dissertation, Indiana University, 1965). Dissertation Abstracts, 1966, 26, 6569-6570. (University Microfilms No. 66-1514)

Wilson, J. W. Evaluation of learning in secondary school mathematics. In B. S. Bloom et al., Handbook on formative and summative evaluation of student learning, 1971.

Footnote

[1]The author is pleased to acknowledge the help of Pamela J. Roelfs in conducting the search of the literature.

TABLE 1

Agreement of Judges in Classifying Items into the Taxonomic Categories

	Number of Judges	Number of Items	Subject Matter	Index of Agreement
Pancella (1970)	12	56 (of 2689)	Biology	84%[a]
Little (1971)	3	148	Geometry	83% agreement
Klein (1972)	3	21 (1 for each subcategory)	Social Studies	Agreement on 16 of the 21 subcategories
Ernst (1974)	10	24 (of approx 1000)[b]	Music	59%[a]
Unruh (1974)	6 (Pooled into 2 groups)	74	Reading	r=.76 for mean item rankings; r=.96, adjusted[c]

Table 1 (Continued)

	Number of Judges	Number of Items	Subject Matter	Index of Agreement
Agishi (1976)	2	228	English	95% agreement
Mills, Rice, Berliner, & Rosseau (1980)	3	Not given; Classroom dialogue	Diverse readings (articles)	Inter-rater r approx. .80

a Agreement with investigator's classification as "model".

b Intended student behaviors from learning activities and evaluation sections of certain teaching units.

c Using Spearman-Brown prophecy formula.